AI OUTLOOK SERIES

23 FEBRUARY 2023 • 10 MINUTE READ

NIST unveils Artificial Intelligence Risk Management Framework

First formal US government guidance on standards in designing, developing, deploying, and using AI systems

Written by: Tony Samp, Daniel Tobey, John Gevertz, Angeline Chen, Andrew Serwin, Matt Dhaiti

The National Institute of Standards and Technology (NIST), an agency of the US Department of Commerce, has released its AI Risk Management Framework (AI RMF) 1.0. With the stated goal of improving the trustworthiness of artificial intelligence, the AI RMF, issued on January 26, provides a structured approach and serves as a "guidance document for voluntary use by organizations designing, developing, deploying or using AI systems to help manage the many risks of AI technologies."

AI's unique nature and challenges underscore the critical role of enterprise risk management practices in ensuring responsible development and use of AI capabilities. This first iteration of the AI RMF is the culmination of over a year and a half of drafting, workshops, and review of hundreds of stakeholder comments. As a stakeholder involved in the process, DLA Piper in October 2021 submitted its "AI Scorebox," a global tool used to determine an organization's AI maturity, to NIST to help inform the development of a framework for managing risks associated with artificial intelligence.

The AI RMF follows by less than three months the issuance of the Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People by the White House Office of Science and Technology Policy (OSTP). The AI Bill of Rights centers on five principles "intended to support the development of policies that protect civil rights and promote democratic values in the building, deployment and governance of automated systems," the document states. (See our alert White House AI Bill of Rights may prompt agency rulemaking and legislation.)

These two Administration initiatives are intended to work together, with the AI RMF serving a functional role offering "AI actors" practical guidance, metrics, and tools to achieve results consistent with the values and principles espoused in the AI Bill of Rights. While intended to be industry agnostic, the AI RMF expresses the need to adapt the framework to existing industry regulations and idiosyncrasies.

Although compliance with the AI RMF is voluntary, the new framework represents an important moment for companies and other organizations looking for information and direction on how to manage AI risks at a time when the regulatory and legislative scrutiny over AI is only bound to increase.

Relevance and framework components

The AI RMF offers a powerful and relevant tool to organizations, equipping them to address the increasingly ubiquitous nature of AI throughout society, multiple industries, and many aspects of organizational activity. As AI technology evolves and become more sophisticated and further integrated into organizational processes and systems, its impact will grow exponentially.

Developing the capability to identify, assess, and manage risks that impact operations, business activities, and objectives ensures that organizations are designed for and operating with optimized efficiency, productivity, and competitiveness. Adopting an integrated approach to enterprise risk management ensures that relevant AI risks are identified and managed in a systematic and consistent manner and enables organizations to become both sustainable and resilient.

The AI RMF adopts fundamental principles of risk management within the context of AI and identifies four "core" functions, with specific actions and outcomes further described for each:

Governance. A risk management culture must be cultivated across the lifecycle of AI systems, including appropriate structures, policies, and processes. Risk management must be a priority for senior leadership, who can set the tone for organizational culture, and for management, who aligns the technical aspects of AI risk management with organizational policies.

Mapping. This function establishes the context to frame risks related to an AI system. Organizations are encouraged to: categorize their AI systems; establish goals, costs, and benefits compared to benchmarks, map risks, and benefits for all components of the AI system; and examine impacts to individuals, groups, communities, organizations, and society.

Measurement. Using quantitative, qualitative, or hybrid risk assessment methods, organizations should analyze AI systems for trustworthy characteristics, social impact, and human-AI configurations.

Management. Identified risks must be managed, prioritizing higher-risk AI systems. Risk monitoring should be applied over time as new and unforeseen contexts, risks, needs, or expectations will emerge.

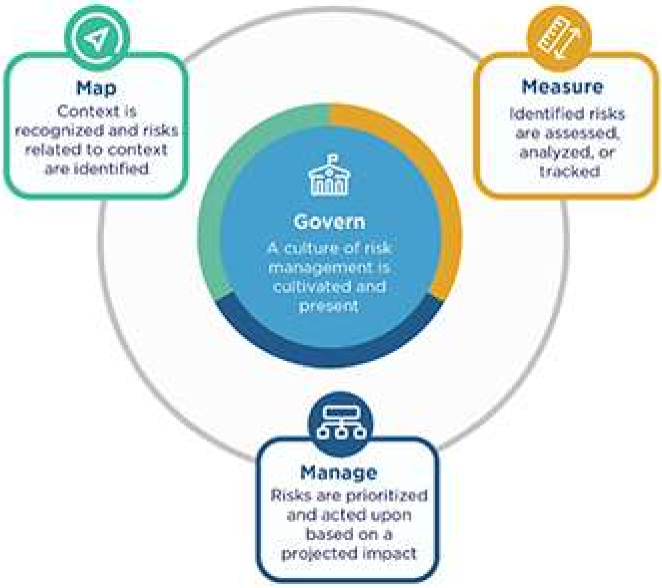
Map
Context is recognized and risks related to context are identified
Measure
Identified risks are assessed, analyzed, or tracked
Govern
A culture of risk management is cultivated and present
Manage
Risks are prioritized and acted upon based on a projected impact

Figure 1 Fig. 5, NIST AI RMF p. 25

Implementation and effective management of these functions ensures that clear communication and actions are established to effectively manage AI risks and ensure development of trustworthy AI systems.

NIST has also published a companion AI Risk Management Framework Playbook as well as several crosswalks (*ie*, tools mapping the AI RMF to other, parallel AI standards) to help users implement the framework. The Playbook is an online resource that will be further updated and is open for feedback or comments until the end of February.

In addition, NIST released an AI Risk Management Framework Roadmap, a list of initiatives for advancing the framework that NIST hopes organizations will carry out independently or in collaboration with the agency.

The AI RMF and the state of AI litigation and enforcement

AI disputes and enforcement are both in their infancy but nonetheless are a growing concern for businesses increasingly reliant upon, or impacted by, AI. Several recent AI-related lawsuits may be seen as bellwethers. These cases highlight a range of liability considerations, among them product liability torts, intellectual property disputes, discrimination claims, and violations of privacy laws.

At the same time, governments and legislatures are recognizing that AI warrants greater scrutiny and oversight, creating new standards of care that affect litigation claims.

NIST reinforces these emerging themes in the AI RMF, contributing to a growing body of literature establishing AI best practices and standards of care. For example, the emphasis on verifying algorithms to ensure they do not propagate harmful bias tracks the surge in enforcement and litigation activity alleging civil rights violations by state and private actors where AI affects access to health, capital, jobs, and other vital resources.

The AI RMF's focus on privacy enhancement underscores the expectation of increased scrutiny of privacy protections considering the vast amounts of data used by AI, as well as AI's ability to reidentify individuals using seemingly anonymized or deidentified data. And NIST's emphasis on the human-machine interface, as well as the many points of human input into AI systems from design to implementation, highlights an emerging trend: the need to identify responsible human actors across the AI lifecycle, particularly at points where the baton is passed (and sometimes dropped) between humans and machines. The RMF identifies 14 unique AI risks that it says are distinct from traditional privacy and cybersecurity issues, raising the bar for an AI-specific standard of care and standardized controls.

The approach taken in the AI RMF reaffirms the trend in global AI regulation that sectors matter; what constitutes bias or fairness in one domain may not work in another. Thus, good internal AI governance must be horizontal *and* vertical, and regulators and litigants are already pressing sector-specific litigation under existing industry standards.

Privacy, security, and AI-specific governance considerations

Through its detailed outline of sound risk management practices and standardization of risk management approaches, the AI RMF provides clear guidance on how to build a strong risk management foundation that enhances an organization's existing governance processes and builds in resiliency for AI. The AI RMF underscores how AI should both benefit from, and add to, existing risk management frameworks for privacy, security, data, and ESG.

AI presents a quintessential big data issue, since nothing requires data of the volume of AI to build, train, and test algorithms. Organizations must ensure that their data governance processes are robust enough to ensure availability and confidence that proper data use rights exist, as well as appropriate controls over the use of data for these purposes. But, as the AI RMF notes, AI poses additional data challenges, including monitoring training data for harmful bias that may infect resulting models, as well as ensuring the representativeness of data sets and ground truths for the ultimate end-user populations. Likewise, on the output side of the equation, risk-management processes must be sufficient to root out harmful bias, model drift, and address other risks raised in the AI RMF. Throughout the life cycle, procedures for explaining the AI and testing the AI for bias and fairness must also be effective and documented.

The “what” of governance, of course, must be balanced by the “who.” The best laid strategies and plans are impotent without effective implementation. The AI RMF provides guidance on addressing necessary accountability in establishing robust AI governance by underscoring the criticality of identifying who in the organization is doing the governance work, ensuring the organization has the requisite skills and resources to implement effective controls.

The AI RMF notes that because AI touches so many parts of an organization, it is imperative that “all AI actors work together to manage risks and achieve the goals of trustworthy and responsible AI” and “are integrated throughout the AI lifecycle.” These diverse stakeholders across an organization provide the varied “technical, societal, legal, and ethical standards or norms” of AI. In many organizations, both the legal and privacy teams are involved because they identify, guide, or set requirements for, many data use issues. Meanwhile, questions of data rights, access, and so on call for effective enterprise data governance that is communicated across the organization, whether that is managed by the privacy team, another group, or in some cases a cross-functional group.

Testing models for bias and fairness requires data science knowledge partnered with legal expertise. Responsibility and execution of these interdependent functions and skills are often found in different parts of the organization. Lawyers should be involved in making sure the processes and testing are effectively documented in ways which, on the one hand, meet increasing regulatory requirements, while, on the other, separate the test process and results from the “special sauce” – the algorithms themselves which are likely to be proprietary.

The legal function must also evaluate whether the processes with regard to both inputs and outputs are adequate to meet regulatory requirements, the organization’s risk appetite, and the level of fairness important to the entity for its use of AI and still comply with an increasingly complex matrix of laws and regulations applicable to the organization’s business processes and activities.

Next steps

Tackling the risks associated with AI is daunting, particular for entities that may be just embarking on evaluations of the impact of AI on their operations or that may not have developed enterprise risk management processes.

The comprehensive and holistic approach presented in the AI RMF can help such organizations consider AI and the associated risks and identify the tools and methods by which such risks can be better managed. For entities already familiar with NIST’s cybersecurity and privacy frameworks and similar processes, the structure of the AI RMF will be familiar and relatively easy to adopt and integrate with existing practices. Even organizations uncertain as to how AI is relevant to their business operations can still benefit from reading the AI RMF and accompanying tools, such as the Playbook and crosswalks.

Ultimately, it remains for each individual organization to determine how to best incorporate the AI RMF into its own risk management and data governance process. As AI becomes more integrated, embedded, and ubiquitous in and across societal and business processes, NIST's voluntary AI RMF provides a powerful and practical set of tools and principles on which to establish a sound AI risk management foundation.

To find out more about the implications of the AI RMF for your organization, please contact the authors.

Related insights

White House AI Bill of Rights may prompt agency rulemaking and legislation

29 NOVEMBER 2022 • 7 MINUTE READ

Your clinical decision support software may now be regulated by FDA as a medical device

29 SEPTEMBER 2022 • 13 MINUTE READ